

Hybrid Maximum Likelihood Modulation Classification Using Multiple Radios

Onur Ozdemir, *Member, IEEE*, Ruoyu Li, and Pramod K. Varshney, *Fellow, IEEE*

Abstract—The performance of a modulation classifier is highly sensitive to channel signal-to-noise ratio (SNR). In this paper, we focus on amplitude-phase modulations and propose a modulation classification framework based on centralized data fusion using multiple radios and the hybrid maximum likelihood (ML) approach. In order to alleviate the computational complexity associated with ML estimation, we adopt the Expectation Maximization (EM) algorithm. Due to SNR diversity, the proposed multi-radio framework provides robustness to channel SNR. Numerical results show the superiority of the proposed approach with respect to single radio approaches as well as to modulation classifiers using moments based estimators.

Index Terms—Modulation classification, data fusion, ML estimation, EM algorithm

I. INTRODUCTION

Modulation classification (MC) is a statistical signal processing problem that deals with determining the modulation type of a noisy communication signal. It plays an important role in many civilian and military applications, e.g., adaptive cognitive radios for satellite communications [1]. It is well known that the optimal classifier (in the Bayesian sense) is the likelihood based (LB) classifier. Different forms of LB classifiers for the MC problem have been proposed in the literature [2]. These include generalized likelihood ratio test (GLRT), average likelihood ratio test (ALRT) and hybrid likelihood ratio test (HLRT) based classifiers. A thorough review of these techniques can be found in [3]. In this paper, we focus on amplitude-phase modulations and consider the HLRT approach, where the likelihood function (LF) is marginalized over the unknown constellation symbols and then the resulting average LF is used to find the ML estimates of the remaining unknown signal parameters. These estimates are then plugged into the average LFs to perform maximum likelihood (ML) classification. We call this approach *hybrid maximum likelihood classification*.

The performance of an MC system using a single radio depends highly on the channel quality, i.e., fading and background noise. In addition, some signal parameters, such as signal-to-noise ratio (SNR) and/or phase offset, are usually unknown which further complicates the classification problem. Note that, for an MC problem, the radio receiver acts as a sensor, therefore, we use the terms radio and sensor interchangeably throughout the paper. Receiver diversity is a

common technique used in wireless communication systems to alleviate channel fading effects for demodulation/symbol detection. Similarly, it is natural to argue that using multiple radios for modulation classification, i.e., collaborative MC, has the potential for improving classification performance compared to a single radio especially in the low to mid signal-to-noise (SNR) regimes. Inspired by this reasoning, collaborative MC approaches have been proposed in [4], [5], [6], [7], [8]. Most of these works are based on the distributed detection framework [9], where each radio makes a local (hard or soft) classification decision and then these decisions are fused at a fusion center (FC) to make a global decision [6], [7], [8]. The only centralized approach proposed in the literature is in [4], where an antenna array is used to receive the unknown signal. The authors use moments based estimators to estimate the unknown signal parameters to simplify the estimation problem. As a result, the estimates in [4] are obtained by ignoring the coupling (due to common received constellation symbols) between different antenna elements which results in sub-optimality.

In this paper, we propose a centralized fusion approach where raw data (instead of decisions) from local radios as in [4] are fused at a fusion center to make the global classification decision. Although the proposed centralized data fusion approach is expected to improve the performance, the resulting MC problem is computationally much harder to solve than a single radio based MC. In order to alleviate this issue, we propose to use the well-known Expectation-Maximization (EM) algorithm [10], which significantly simplifies the MC problem along with its nice convergence properties. In an earlier work [11], the EM algorithm was used for the MC problem using a single radio under flat fading channels corrupted by Gaussian mixture noise. Our proposed framework along with the problem formulation for centralized fusion based MC is different from the problem considered in [11] even though the EM algorithm is suitable for both. Due to SNR diversity, the proposed centralized data fusion framework significantly improves the MC performance compared to single radio approaches as in [11]. Furthermore, our numerical results show that the proposed EM based solution provides superior performance compared to the moments based solution proposed in [4] with only a small increase in computational complexity.

II. PROBLEM FORMULATION

Consider a radio/sensor network with L sensors observing the same communication signal with a block of N constellation (information) symbols that undergo flat fading. These

O. Ozdemir is with ANDRO Computational Solutions, 7902 Turin Road, Rome, NY 13440. R. Li is with the School of EEE, Nanyang Technological University, Singapore. P. K. Varshney is with Department of EECS, Syracuse University, Syracuse, NY 13244. This material is in part based upon work supported by the Government under Contract No. W15P7T-12-C-A040. Email: oozdemir@androcs.com, rli11@syr.edu, varshney@syr.edu.

sensors are located far enough from each other such that they experience independent fading. We assume that timing and frequency offsets have been perfectly estimated. The received baseband observation sequence at sensor l can be written as:

$$r_{l,n} = a_l e^{j\theta_l} I_n + w_n, \quad (1)$$

where $l = 1, \dots, L$, $n = 0, \dots, N-1$, I_n is the n^{th} complex constellation symbol of the block, w_n is the additive complex zero-mean white Gaussian noise with variance N_0 , and a_l and θ_l are the channel gain and the channel phase at sensor l , respectively. The above signal model is a commonly used model in the MC literature [2], [11], [12], [13]. In this model, $\{a_l\}_{l=1}^L$, $\{\theta_l\}_{l=1}^L$, $\{I_n\}_{n=0}^{N-1}$ are the unknown signal parameters. In a general modulation classification scenario, in addition to the unknown signal parameters, noise power N_0 may also be unknown. In this case, the unknown parameter vector can be expressed as $\tilde{\mathbf{u}} := [\mathbf{a}, \theta, \mathbf{I}, N_0]$, where $\mathbf{a} := [a_1, \dots, a_L]^T$, $\theta := [\theta_1, \dots, \theta_L]^T$ and $\mathbf{I} := [I_0, \dots, I_{N-1}]^T$. We assume that noise is independent across sensors. Suppose there are S candidate modulation formats under consideration and let $I_n^{(i)}$ denote the constellation symbol at time n corresponding to modulation $i \in \{1, \dots, S\}$. Let \mathbf{r} denote the observation vector defined as $\mathbf{r} := [\mathbf{r}_1^T, \dots, \mathbf{r}_L^T]^T$ where $\mathbf{r}_l := [r_{l,0}, \dots, r_{l,N-1}]^T$ and H_i represents the hypothesis associated with modulation format i . Let $p_i(\mathbf{r}|\tilde{\mathbf{u}}) := p(\mathbf{r}|H_i, \tilde{\mathbf{u}})^2$ denote the conditional probability density function (pdf) of \mathbf{r} conditioned on the unknown modulation format i and the unknown parameter vector $\tilde{\mathbf{u}}$, i.e., the likelihood function (LF), which is given by

$$p_i(\mathbf{r}|\tilde{\mathbf{u}}) = \frac{1}{(\pi N_0)^{NL}} \exp \left(-\frac{1}{N_0} \sum_{l=1}^L \sum_{n=0}^{N-1} |r_{l,n} - a_l e^{j\theta_l} I_n^{(i)}|^2 \right). \quad (2)$$

Note that the LF in (2) is parameterized by the constellation symbols $I_n^{(i)}$, which represent the modulation format i . This is a composite multiple hypothesis testing problem. In a Bayesian setting, the optimal classifier in terms of minimum probability of classification error is the maximum a posteriori (MAP) classifier. If there is no information available on *a priori* probabilities, which is usually the case in a noncooperative environment, one can use a non-informative prior, i.e., each modulation scheme is assigned an identical prior probability. This is the assumed scenario in which case the optimal classifier takes the form of a maximum likelihood (ML) classifier.

In the hybrid maximum likelihood approach, the LF is averaged over the unknown constellation symbols I_n and then maximized over the remaining unknown parameters. Let $\mathbf{u} := [\mathbf{a}, \theta, N_0]$. Given I_n and modulation format i , we have the following

$$p_i(r_{1,n}, \dots, r_{L,n} | I_n, \mathbf{u}) = \prod_{l=1}^L p_i(r_{l,n} | I_n, \mathbf{u}). \quad (3)$$

After averaging over I_n ,

$$p_i(r_{1,n}, \dots, r_{L,n} | \mathbf{u}) = \frac{1}{M_i} \sum_{m=1}^{M_i} \prod_{l=1}^L p_i(r_{l,n} | I_n^{m,(i)}, \mathbf{u}), \quad (4)$$

where M_i and $I_n^{m,(i)}$ are the number of constellation symbols and the m^{th} constellation symbol in modulation i , respectively. For example, $M_i = 2$ for BPSK since BPSK has only two possible constellation symbols. Note that, in (4), the constellation symbols are assumed to have equal *a priori* probabilities, i.e., $p(I_n^{m,(i)} | H_i) = 1/M_i$. This is a common assumption made in communication theory. Without loss of generality, we further assume that $\mathbb{E}\{|I_n^{(i)}|^2\} = 1$, where $\mathbb{E}\{\cdot\}$ denotes statistical expectation. In other words, the power of constellation symbols is normalized to unity. Using (4), $p_i(\mathbf{r}|\mathbf{u})$ becomes

$$p_i(\mathbf{r}|\mathbf{u}) = \frac{1}{M_i^N (\pi N_0)^{NL}} \prod_{n=0}^{N-1} \sum_{m=1}^{M_i} \exp \left(-\frac{1}{N_0} \sum_{l=1}^L |r_{l,n} - a_l e^{j\theta_l} I_n^{m,(i)}|^2 \right). \quad (5)$$

Taking the natural logarithm of the above LF and discarding constant terms, we have the log-likelihood function (LLF) in (6) shown on the top of next page. In HLRT, the modulation format that maximizes the resulting LLF is selected as the final decision, i.e.,

$$\hat{i} = \arg \max_i \Lambda_i(\hat{\mathbf{u}}_i), \quad (7)$$

where

$$\hat{\mathbf{u}}_i = \arg \max_{\mathbf{u}} \Lambda_i(\mathbf{u}). \quad (8)$$

From (6), we can make the following observations. The problem of finding the global maximum of $\Lambda_i(\mathbf{u})$ with respect to \mathbf{u} is a $2L+1$ dimensional non-convex optimization problem which is extremely difficult to solve in general. Furthermore, there is coupling between the unknowns of different sensors due to common unknown constellation symbols. In other words, the problem cannot be decoupled into equivalent multiple lower dimensional (simpler) optimization problems. There is no closed-form analytical solution. Therefore, either numerical methods or approximation techniques need to be employed. In the following section, we discuss our approach for solving this problem which is based on the Expectation-Maximization (EM) algorithm.

III. THE EM ALGORITHM

Suppose modulation format i is under consideration and the constellation symbol vector \mathbf{I} is known. In this case, we have the following closed-form expressions for the ML estimators:

$$\hat{\theta}_l = \tan^{-1} \left(\frac{\Im(\mathbf{I}^H \mathbf{r}_l)}{\Re(\mathbf{I}^H \mathbf{r}_l)} \right), \quad (9)$$

$$\hat{a}_l = \frac{\Re \left(e^{-j\hat{\theta}_l} \mathbf{I}^H \mathbf{r}_l \right)}{\sum_{n=0}^{N-1} |I_n|^2}, \quad (10)$$

$$\hat{N}_0 = \frac{1}{LN} \sum_{n=0}^{N-1} \sum_{l=1}^L |r_{l,n} - \hat{a}_l e^{j\hat{\theta}_l} I_n|^2, \quad (11)$$

¹Superscript T denotes vector/matrix transpose.

²Throughout the paper, we use the notation $p_i(\cdot)$ to denote $p(\cdot|H_i)$.

$$\Lambda_i(\mathbf{u}) = -LN \ln N_0 + \sum_{n=0}^{N-1} \ln \left(\sum_{m=1}^{M_i} \exp \left(-\frac{1}{N_0} \sum_{l=1}^L |r_{l,n} - a_l e^{j\theta_l} I_n^{m,(i)}|^2 \right) \right) \quad (6)$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote real and imaginary parts of a complex number, respectively, and H denotes the Hermitian of a complex vector/matrix. From the above closed-form expressions, it is clear that when \mathbf{I} is known, the maximization problem (for estimating a_l and θ_l) is decoupled between different sensors. Due to the fact that the ML estimation problem is significantly simpler when the constellation symbols are known, we adopt the well-known EM algorithm [10] to solve this problem. The EM algorithm is an iterative method which enables the computation of ML estimates, especially well suited to problems where ML estimation is intractable because of the presence of unknown (unobserved) data. In our case, the constellation symbols represent unobserved data. We can formally describe the EM algorithm for our problem in (8) as follows [10]. Let us define the so-called *complete data* $\mathbf{x} := [\mathbf{r}^T, \mathbf{I}^T]$. Starting from an initial estimate $\hat{\mathbf{u}}_i^{(0)}$, the EM algorithm performs the following two steps: the expectation step (E-step) and the maximization step (M-step).

$$\textbf{E-step: } Q(\mathbf{u}_i | \hat{\mathbf{u}}_i^{(t)}) = \mathbb{E} \left\{ \ln p_i(\mathbf{x} | \mathbf{u}_i) | \mathbf{r}, \hat{\mathbf{u}}_i^{(t)} \right\}, \quad (12)$$

$$\textbf{M-step: } \hat{\mathbf{u}}_i^{(t+1)} = \arg \max_{\mathbf{u}_i} Q(\mathbf{u}_i | \hat{\mathbf{u}}_i^{(t)}). \quad (13)$$

Given the fact that the unknown parameter vector \mathbf{u} is independent of the transmitted constellation symbols \mathbf{I} , the E-step in (12) reduces to:

$$Q(\mathbf{u}_i | \hat{\mathbf{u}}_i^{(t)}) = \sum_{\mathbf{I}} \ln p_i(\mathbf{r} | \mathbf{I}, \mathbf{u}_i) P_i(\mathbf{I} | \mathbf{r}, \hat{\mathbf{u}}_i^{(t)}). \quad (14)$$

It is straightforward to calculate $\ln p_i(\mathbf{r} | \mathbf{I}, \mathbf{u}_i)$ using (2). Let $\mathbf{r}_n := [r_{1,n}, \dots, r_{L,n}]^T$ and $\alpha_n^{m,(t)} := P_i(I_n = I^m | \mathbf{r}_n, \hat{\mathbf{u}}_i^{(t)})$, $m = 1, \dots, M_i$, denote the *a posteriori* probability of the unknown constellation symbol which can be calculated as

$$\begin{aligned} \alpha_n^{m,(t)} &:= P_i(I_n = I^m | \mathbf{r}_n, \hat{\mathbf{u}}_i^{(t)}) = \frac{p_i(I_n = I^m, \mathbf{r}_n | \hat{\mathbf{u}}_i^{(t)})}{P_i(\mathbf{r}_n | \hat{\mathbf{u}}_i^{(t)})} \\ &= \frac{p_i(\mathbf{r}_n | I_n = I^m, \hat{\mathbf{u}}_i^{(t)})}{\sum_{k=1}^{M_i} p_i(\mathbf{r}_n | I_n = I^k, \hat{\mathbf{u}}_i^{(t)})}. \end{aligned} \quad (15)$$

In (15), we have used the assumption that each data symbol has the same *a priori* probability, i.e., $P_i(I_n = I^m | \hat{\mathbf{u}}_i^{(t)}) = 1/M_i$, $m = 1, \dots, M_i$. Let us also define

$$v_n^{(t)} := \sum_{m=1}^{M_i} \alpha_n^{m,(t)} I^m, \quad E^{(t)} := \sum_{n=0}^{N-1} \sum_{m=1}^{M_i} \alpha_n^{m,(t)} |I_n^m|^2. \quad (16)$$

Note that $v_n^{(t)}$ and $E^{(t)}$ represent the *a posteriori* expectations of the constellation symbol at time n and the total normalized energy of the transmitted discrete-time signal, respectively. Substituting (15)-(16) in (14) and carrying out the maximization in (13) by taking the first derivatives and setting them to

zero, we obtain the following closed form expressions for the $(t+1)$ -th step in the EM algorithm:

$$\hat{\theta}_l^{(t+1)} = \tan^{-1} \left(\frac{\Im(\mathbf{\Upsilon}^{(t)H} \mathbf{r}_l)}{\Re(\mathbf{\Upsilon}^{(t)H} \mathbf{r}_l)} \right), \quad (17)$$

$$\hat{a}_l^{(t+1)} = \frac{1}{E^{(t)}} \Re \left(e^{-j\hat{\theta}_l^{(t+1)}} \mathbf{\Upsilon}^{(t)H} \mathbf{r}_l \right), \quad (18)$$

$$\hat{N}_0^{(t+1)} = \frac{1}{LN} \sum_{n=0}^{N-1} \sum_{m=1}^{M_i} \alpha_n^{m,(t)} \sum_{l=1}^L |r_{l,n} - \hat{a}_l^{(t+1)} e^{j\hat{\theta}_l^{(t+1)}} I_n^m|^2, \quad (19)$$

where $\mathbf{\Upsilon}^{(t)} := [v_0^{(t)}, \dots, v_{N-1}^{(t)}]^T$. One important property of the EM algorithm is that the LF monotonically increases at every iteration and converges to a stationary point [14]. However, this stationary point can be a local maxima, therefore, either a good initialization or multiple initializations are needed to guarantee convergence to a good stationary point.

IV. METHOD OF MOMENTS ESTIMATORS AND EM INITIALIZATION

In the literature, there have been attempts to develop simple estimators for \mathbf{u} due to the complexity associated with the ML estimator in the MC problem. These estimators are based on the method of moments (MoM) [4], [15]. The MoM estimators for a_l and N_0 are given, respectively, as [4], [15]:

$$\hat{a}_{l,(i)} = \left(\frac{2\hat{M}_{2,l}^2 - \hat{M}_{4,l}}{2 - \mathbb{E}\{|I|^4\}} \right)^{1/4}, \quad (20)$$

$$\hat{N}_{0(i)} = \sum_{l=1}^L \frac{\hat{a}_{l,(i)} \hat{N}_{0l,(i)}}{\sum_{k=1}^L \hat{a}_{k,(i)}}, \quad (21)$$

where $\hat{N}_{0l,(i)} = \hat{M}_{2,l} - \hat{a}_{l,(i)}^2$; and $M_{2,l} = \mathbb{E}\{|r_{l,n}|^2\}$ and $M_{4,l} = \mathbb{E}\{|r_{l,n}|^4\}$ represent second and fourth absolute moments of $r_{l,n}$, respectively. Note that we use a weighted average to compute $\hat{N}_{0(i)}$ by using the noise power estimate at each sensor, $N_{0l,(i)}$, weighted by its corresponding channel amplitude $\hat{a}_{l,(i)}$. The estimates of the second and fourth absolute moments are given as $\hat{M}_{2,l} = N^{-1} \sum_{n=0}^{N-1} |r_{l,n}|^2$ and $\hat{M}_{4,l} = N^{-1} \sum_{n=0}^{N-1} |r_{l,n}|^4$. Regarding the channel phases, the MoM estimators depend on the modulation format under consideration. If modulation i is M-PSK, the MoM estimate is $\theta_{l,(i)} = M_i^{-1} \arg \left(\sum_{n=0}^{N-1} r_{l,n}^{M_i} \right)$. Otherwise, if modulation i is M-QAM, $\theta_{l,(i)} = \frac{1}{4} \arg \left(\sum_{n=0}^{N-1} r_{l,n}^4 \right)$. The MoM estimates have been used in the literature [4], [13] to replace the ML estimates in (7). These modulation classifiers have been named Quasi HLRT or QHLRT [2], [4], [13]. Here, we propose to use the MoM estimators as initial points ($\mathbf{u}_i^{(0)}$) for the EM algorithm explained in Section III. The EM based ML classification is expected to improve the performance over QHLRT. However, it clearly requires increased computational complexity. Nevertheless, it is still computationally much simpler to implement than numerically searching for the global ML estimates using (6).

V. NUMERICAL RESULTS

In this section, numerical results are provided to show the effectiveness of the proposed EM based solution for MC problem. We consider a 3-ary MC scenario where the modulations under consideration are QPSK, 8-PSK and 16-PSK. The channels are modeled as Rayleigh fading channels, i.e., A is a Rayleigh distributed random variable with scale parameter σ , where the channel SNR is $\mathbb{E}\{A^2\}/N_0 = 2\sigma^2/N_0$ with fixed $N_0 = 1$. Fig. 1 shows the probability of correct classification (P_c) versus channel SNR for different number of radios. The number of samples is fixed at $N = 500$. For comparison, we include the results obtained by MoM estimators as in [4]. Low to mid channel SNR regimes are considered as this is where the multi-radio approach is expected to provide significant performance improvement. First, it is clear from Fig. 1 that a centralized data fusion based multi-radio approach is the key to improving performance at low to mid SNR regimes. For example, at $P_c \approx 0.8$, an SNR gain of 15 dB is attained with 10 radios compared to a single radio. Second, the proposed EM based centralized ML estimation provides superior performance compared to MoM based estimation. In fact, it is surprising to see that the performance of classifiers using MoM based estimation degrades as the number of radios increases to the point where they are no better than simple guessing. This is due to the fact that MoM estimators do not always provide meaningful estimates and they do not necessarily maximize the LF. Moreover, MoM estimators do not take into account coupling between estimates of different radio signals due to common constellation symbols. These factors result in poor sub-optimality of MoM based modulation classifiers. In Fig. 2, the channel SNR is fixed at 0 dB and the performance is depicted with respect to sample size (N). The results are similar to those in Fig. 1. As N increases, P_c also increases. However, it is clear from Fig. 2 that it is better to increase the number of radios than to increase the number of samples. This is due to the SNR diversity gained by employing multiple radios and the fact that each radio experiences flat fading. The trade-off here is the cost of the radios as well as synchronization overhead that is needed between radios.

VI. CONCLUSIONS

We proposed a centralized data fusion based modulation classification framework using multiple radios and the hybrid ML approach. Our proposed solution is based on the EM algorithm which significantly simplifies the complicated ML estimation problem. Numerical results show that the proposed approach is superior to single radio approaches as well as classifiers using moments based estimators.

REFERENCES

- [1] J. Hamkins, M. K. Simon, and J. H. Yuhen, *Autonomous Software-Defined Radio Receivers for Deep Space Applications (JPL Deep-Space Communications and Navigation Series)*. Wiley-Interscience, 2006.
- [2] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET Communications*, vol. 1, no. 2, pp. 137–159, Apr. 2007.
- [3] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Trans. Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 41, no. 4, pp. 455–469, Jul. 2011.

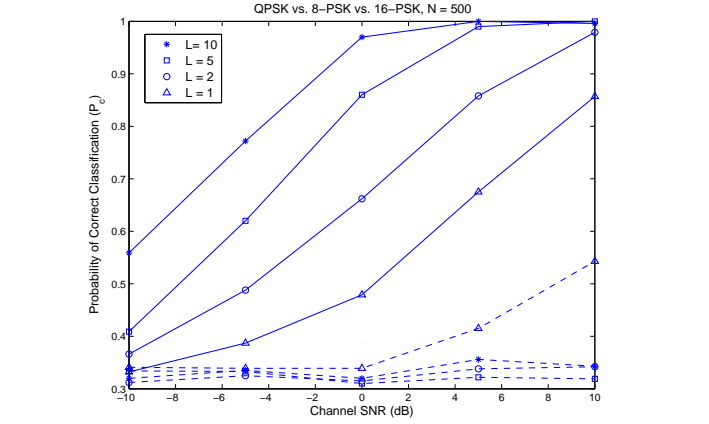


Fig. 1. Probability of correct classification versus channel SNR. Solid lines: EM based ML estimation, dashed lines: MoM estimation.

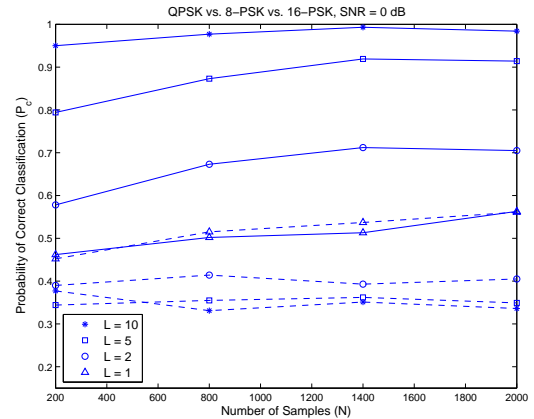


Fig. 2. Probability of correct classification versus number of samples. Solid lines: EM based ML estimation, dashed lines: MoM estimation.

- [4] A. Abdi, O. A. Dobre, R. Chauchy, Y. Bar-Ness, and W. Su, "Modulation classification in fading channels using antenna arrays," in *Proc. IEEE MILCOM*, Monterey, CA, Nov. 2004, pp. 211–217.
- [5] W. Su and J. Kosinski, "Framework of network centric signal sensing for automatic modulation classification," in *Proc. IEEE ICNSC*, Chicago, IL, Apr. 2010, pp. 534–539.
- [6] J. L. Xu, W. Su, and M. Zhou, "Distributed automatic modulation classification with multiple sensors," *IEEE Sensors Journal*, vol. 10, no. 11, pp. 1779–1785, Nov. 2010.
- [7] —, "Asynchronous and high-accuracy digital modulated signal detection by sensor networks," in *Proc. IEEE Military Communications Conf. (MILCOM)*, Nov. 2011.
- [8] Y. Zhang, N. Ansari, and W. Su, "Optimal decision fusion based automatic modulation classification by using wireless sensor networks in multipath fading channel," in *Proc. IEEE Global Communications Conf. (GLOBECOM)*, Dec. 2011.
- [9] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1997.
- [10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal Roy. Stat. Soc. (Series B)*, vol. 39, no. 1, pp. 1–38, 1977.
- [11] V. G. Chavali and C. R. C. M. da Silva, "Maximum-likelihood classification of digital amplitude-phase modulated signals in flat fading non-Gaussian channels," *IEEE Trans. Communications*, vol. 59, no. 8, pp. 2051–2056, Aug. 2011.
- [12] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "On the classification of linearly modulated signals in fading channels," in *Proc. Conf. on Information Sciences and Systems (CISS)*, Mar. 2004.
- [13] F. Hameed, O. A. Dobre, and D. C. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Comm.*, vol. 8, no. 12, pp. 5884–5892, Dec. 2009.
- [14] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983.
- [15] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum, 1997.